

Predicting Gait Retraining Strategies for Knee Osteoarthritis

Benjamin Wittevrongel^{1*}, Irma Ravkic^{2*}, Wannes Meert², Jesse Davis², Tim Gerbrands^{3,4} and Benedicte Vanwanseele^{3,4}

¹ Laboratory for Neuro- and Psychophysiology, KU Leuven, Belgium

² Departement of Computer Science, KU Leuven, Leuven, Belgium

³ Departement of Kinesiology, KU Leuven, Leuven, Belgium

⁴ Health Innovation and Technology, Fontys University of Applied Sciences

Abstract. Symptomatic knee osteoarthritis is one of the most common types of osteoarthritis and is one of the top ten causes of years of life lost due to disability. One common treatment strategy involves the prolonged use of anti-inflammatory medications. Recently, gait retraining has been proposed as non-invasive and non-pharmacological treatment option for knee osteoarthritis. However, many possible gait retraining strategies exist and it is unknown a priori which strategy will work best for a given patient. Hence, it is often necessary to pursue a trial-and-error approach to find the best strategy. In this paper we investigate using two standard machine learning techniques, decision trees and rule sets, to build models based on features of a subject’s normal gait to predict which strategy will work best. We were able to learn reasonably accurate models for this task. Furthermore, as the learned models are interpretable, a domain expert was able to gain several insights into the problem by inspecting them. This work shows that machine learning can be useful for prediction of treatment strategies in rheumatology.

Keywords: knee osteoarthritis, gait retraining prediction, machine learning

1 Introduction

Symptomatic knee osteoarthritis (OA) is one of the most common types of OA and is one of the top ten causes of ‘years of life lost due to disability’ [7]. Approximately 44% of the human population will develop symptoms of this disease during their lifetime [11] and this percentage rises to 66% among people suffering from obesity [11]. Additionally, OA entails high socio-economic costs [2, 1], mainly due to visits to doctors, surgery, and medication as well as indirect costs such as missed work time.

Gait retraining has recently been proposed as a non-invasive treatment strategy for knee OA because several studies have shown that a high external knee adduction moment (EKAM) during walking is closely related to the progression

* These authors contributed equally to this work.

of knee OA as the EKAM reflects the medio-lateral load distribution on the tibio-femoral joint [9, 13]. Gait retraining attempts to slow OA’s progression by modifying a patient’s gait kinematics to reduce the EKAM. Several gait modification techniques have been tested and shown to effectively reduce the EKAM, including medialising the knee during the stance phase (*Medial Thrust*) [3], leaning the trunk in the direction of the stance leg (*Trunk Lean*) [10] and increasing the toe-out angle [12], among others.

Determining the best gait retraining strategy in practice usually relies on intuition and experimentation which can be imprecise and time consuming. In contrast, Fregly et al. [4, 3] designed the Medial Thrust retraining strategy by using a computational approach. They employed a dynamic optimization of a patient-specific, full-body gait model to predict the 3-D gait modifications that reduce both peaks of the external knee adduction torques. While this strategy was able to pinpoint exactly how a patient can achieve the greatest EKAM reduction, calculating full-body models necessitated the use of lab expensive motion capture camera system to capture the patient’s gait as well as an expensive software package. Additionally, the study was performed on only one subject.

In this paper, we will tackle the task of selecting the appropriate gait retraining strategy as a machine learning problem. Specifically, we will look at building a model based on characteristics of an individual’s habitual gait. We will examine data from both healthy and arthritic subjects that was collected as part of a study to compare which strategies are most effective in practice [5]. We obtained a model with an area under the ROC curve of 0.75 when training on healthy subjects and 0.92 when training on arthritic subjects. Additionally, we presented the learned models to a domain expert for interpretation. She determined a promising location for attaching a gyroscope to an individual in order to measure gait characteristics that are predictive of the best retraining strategy. This offers the potential to avoid using an expensive motion capture camera to select a retraining strategy.

2 Materials and Methods

This section describes the data set we used, challenges we encountered processing the data, and our experimental methodology for predicting the best gait retraining strategy.

2.1 Data Collection

The data we use has been collected as part of Tim Gerbrands’ PhD thesis. The study includes 62 subjects aged between 18 and 65 of which 28 subjects were diagnosed with symptomatic medial tibiofemoral knee osteoarthritis and the rest were healthy individuals. Each subject is described by a number of gait features and the expert determined the class label to be the gait retraining strategy that maximally reduced the external knee adduction moment. The subset of this data containing only healthy patients is analysed in the work of Gerbrands et al. [5].

The data was collected as follows. In order to acclimate to the environment and equipment, each participant walked freely and barefoot on a 13 meter long walkway. Next, each subject was asked to perform the three walking conditions described in Table 1 and an investigator provided visual examples for each condition. Each participant was instructed to implement the strategies to the greatest extent possible at a self-selected speed such that walking was still comfortable. A practice period of five minutes was allowed, during which the investigator provided verbal feedback. Each subject performed the normal walking condition first and then the remaining conditions were presented in a randomized order. Five successful trials of each condition were captured. Between conditions, a subject was asked to walk comfortably for approximately three minutes in order to minimize interference from one condition to the next. For each subject, the most effective gait modification was determined to be the one that resulted in the largest positive reduction in external knee adduction moment compared to the normal walking condition.

Walking condition	Instruction
Normal Walking	Walk freely and comfortably as you would on the street.
Trunk Lean	Lean right with the torso as the right foot has floor contact.
Medial Thrust	Move the right knee inwards/medial during right legged stance.

Table 1: Instructions given to the subjects for each walking condition.

During each trial, a variety of sensors are employed to measure various aspects of the subject’s gait. Joint angles of the leg and torso were measured at a frequency of 100Hz with a dual camera wireless active 3D-system (Charnwood Dynamics Ltd., Codamotion CX 1). Ground reaction force was measured at a frequency of 1000Hz during one step per trial with a recessed force plate (Advanced Mechanical Technology, Inc., OR 6-7). EKAM was calculated through inverse dynamics in which the knee centre served as the shank-fixed axes’ origin. Each subject is then described by eleven gait features which are listed in Table 2.

Gerbrands et al. [5] showed that the Medial Thrust and Trunk Lean strategies affect both the overall peak and impulse and also provide the greatest reduction in EKAM. The former reduces the EKAM by bringing the knee joint closer to the center axis of the body and the latter reduces EKAM by having the patient lean more towards the supporting leg. Of the 34 healthy subjects, 18 were assigned *Medial Thrust* as the class label and 16 have *Trunk Lean*. Of the 28 arthritic subjects, four have *Medial Thrust* as the class label and 24 have *Trunk Lean*.

Feature	Description
Knee Flexion	The maximal extent to which the subject bends the knee over one gait cycle.
Trunk Angle	The angle of the upper body while supporting on the arthritic leg.
Knee Adduction	Minimal deviation from perfect alignment of upper and lower leg over one gait cycle.
Knee Abduction	Maximal deviation from perfect alignment of upper and lower leg over one gait cycle.
Tibia Angle	Defines the maximal deviation of the tibia from the perpendicular angle to the ground.
Toe Out Angle	The extent to which the subject moves the toes outwards during gait.
Knee Adduction Moment Peak	Maximal absolute value of the EKAM over one gait cycle.
Knee Adduction Moment Impulse	Area under the EKAM curve over one gait cycle.
1st Peak vGRF	The first peak in the vertical component of the ground reaction force.
2nd Peak vGRF	The second peak in the vertical component of the ground reaction force.
Walking Speed	The speed at which the subject walked during the trials.

Table 2: Gait features used in classification task.

2.2 Data Challenges and Preprocessing

Before applying the machine learning approaches, we performed an iterative investigation of the data to ensure its consistency. First, we identified outlier values by looking at histograms and boxplots for each feature. We identified one value of toe-out angle that the domain expert determined was an outlier, and we marked this value as missing. Next, based on the assumption that the same type of movement for both patient classes should have similar characteristics, we calculated the mean, minimum and maximum value for each feature in the data. We observed in the data that the value of the maximal trunk angle for some subjects was lower than their minimal trunk angle. Finally, several issues arose when comparing the data from healthy and arthritic subjects, which were collected at different times. One was that identical features were coded with different names, which we resolved by consulting with the domain expert. A subtler issue was that several features had different signs. This arose because computing them required choosing one leg as a reference. All healthy subjects used the same leg, whereas for the arthritic subjects the selected leg depended on which knee was arthritic. Discovering this required several iterations with the domain experts.

Nine out of 11 features in the data set had at least one and at most three values missing. For each feature we replace its missing values with the average value of the known values of the feature. Since machine learning techniques

can be sensitive to the range of values that numeric attributes take on [15], we explored whether discretization could improve the results. In this paper we used k -equal frequency binning which divides the data into k groups such that each group contains approximately the same number of values. When handling missing values and discretization, we only considered the training data when computing the average value of a missing feature and selecting the bin widths for discretization.

2.3 Methodology

We use the following three versions of the data in our experiments:

1. **Healthy.** This subset of the data contains individuals who do not suffer from arthritis.
2. **Arthritic.** This subset of the data consists of individuals who have knee osteoarthritis.
3. **Combined.** This data set combines both the healthy and arthritic subjects.

For each of these data sets we will use the classifiers and empirical evaluation as specified below.

Classifiers In the experiments, we compare the predictive power of two classification models: decision trees and rule sets [8]. These learners both produce models that are easily understandable, which enables domain experts to analyze the learned models without needing in-depth knowledge of machine learning. A decision tree consists of internal nodes which represent tests performed on attributes and leaf-nodes which decide the label of an instance. Decision trees classify instances by sorting them from the root of the tree to a leaf-node reached by following the path established by successful internal node tests. A rule set classifier makes a prediction based on a set of IF-THEN rules. The IF part of a rule contains a number of tests on features. If the IF part is satisfied for some instance, the THEN part determines its label. The rules are applied and evaluated in order from first to last. If none of the IF-THEN rules are satisfied, a default label is assigned.

The implementations of the decision tree and rule set learner we use are provided by Weka [6] and we use the J48 [14, 15] tree learning algorithm and the PART [15] rule learner. We configured both algorithms to not allow pruning. While this decision risks the overfitting, we hope to learn more informative models that provide insight into the underlying process. Furthermore, we fixed the minimum number of instances each leaf must contain to three.

Evaluation Methodology We perform leave-one-patient out cross-validation to estimate the predictive performance of the learned models because this evaluation methodology is typically used when there are very few examples. This means that we repeatedly learn models on all but one patient, and use these

models to predict the best retraining strategy for the left-out patient. We evaluate our models by reporting their prediction accuracy. Because the accuracy is influenced by the skewed class distribution, we indicate which results outperform the baseline classifier that always predicts the most frequent class label in the training data. We also report the area under the ROC curve (AUC) for the best model for each data set.

3 Experiments

The goal of the experiments is to explore and answer the following questions:

- Q1.** Can we learn accurate models for predicting the best gait retraining strategy?
- Q2.** Will the same model apply to both the healthy and arthritic populations?
- Q3.** Do the learned models provide insights for a domain expert?
- Q4.** Can the learned models provide guidance for placing an accelerometer or a gyroscope on an individual to enable the relevant measurements to be made outside of an expensive lab set up?

3.1 Results

Next, we present experimental results for data of healthy, arthritic, and combined data set.

Healthy. Table 3 presents the results on the healthy patients. This appears to be a hard prediction problem as many learned models do not outperform the majority classifier. However, when discretizing into 5-equal frequency bins we are able to learn reasonably accurate models.

Discretization	Classifier			
	DECISION TREE		RULE SET	
	Accuracy	AUC	Accuracy	AUC
<i>None</i>	52.9%	0.38	44.1%	0.52
<i>3-equal frequency binning</i>	41.2%	0.38	44.1%	0.52
<i>4-equal frequency binning</i>	47.1%	0.54	47.1%	0.52
<i>5-equal frequency binning</i>	61.8%†	0.52	76.5%†	0.75

Table 3: The accuracies and AUCs for the models learned using only the non-arthritic subjects. The best result is in bold, and † denotes results that are better than majority classifier which has accuracy of 52.9% for this data set.

Arthritic. Table 4 presents the results when training only on arthritic subjects. Again, we see that this is a challenging problem and the only model that outperforms the majority classifier is a decision tree built on data discretized into 5-equal frequency bins.

Discretization	Classifier			
	DECISION TREE		RULE SET	
	Accuracy	AUC	Accuracy	AUC
<i>None</i>	85.7%	0.72	85.7%	0.72
<i>3-equal frequency binning</i>	78.6%	0.67	85.7%	0.68
<i>4-equal frequency binning</i>	75.0%	0.51	67.9%	0.46
<i>5-equal frequency binning</i>	92.9%†	0.92	85.7%	0.80

Table 4: The accuracies and AUCs for the models trained using only the arthritic subjects. The best result is in bold, and † denotes results that are better than majority classifier which has accuracy of 85.7% for this data set.

Combined. Table 5 presents results learned on the combined data set of healthy and arthritic subjects. None of the models learned outperforms the majority classifier.

One explanation is that arthritic subjects have changed their habitual gait to cope with pain and joint immobility associated with the disease. To further explore this hypothesis, we used the best learned model from the healthy subjects to predict the gait retraining strategy for each arthritic subject. This resulted in an accuracy of 75.0%, which is worse than the majority classifier. Similarly, we used the best model learned from arthritic subjects to predict the gait retraining strategies for each healthy subject. Again, this performed worse than the majority classifier with an accuracy of 47.1% and AUC of 0.48. This provides some additional evidence that there are differences in the habitual gait features for the healthy and arthritic subjects.

Discretization	Classifier			
	DECISION TREE		RULE SET	
	Accuracy	AUC	Accuracy	AUC
<i>None</i>	50.0%	0.31	32.25%	0.32
<i>3-equal frequency binning</i>	58.1%	0.60	61.3%	0.61
<i>4-equal frequency binning</i>	51.6%	0.38	54.8%	0.56
<i>5-equal frequency binning</i>	45.2%	0.43	59.7%	0.59

Table 5: Accuracies and AUCs for the decision trees and rule sets learned on data consisting of both healthy and arthritic subjects. The best result is in bold, and † denotes results that are better than majority classifier which has accuracy of 64.5% for this data set.

3.2 Discovered Knowledge

We were also interested if the learned models for healthy and arthritic patients could provide any domain insight. Using the best determined settings, we learned

a decision tree and a rule set from each data set where we used all available examples for training the model. We presented the resulting models to a knee biomechanics and gait retraining expert for interpretation.

Healthy. The best decision tree learned for healthy patients is shown in Figure 1. The expert identified the most interesting branch in the tree as the one with the narrow ranges of values for knee abduction ($-4.0, -3.0]$ and adduction ($0.8, 2.3]$ that results in a prediction of *Trunk Lean*. She theorized that if there is not a lot of movement in the knees during the normal gait then the subjects are better with the *Trunk Lean* strategy.

The best rule set for healthy patients is depicted in Figure 2. The subjects for whom the first rule applies already bring their knee inwards a lot during their natural gait and thus already implement *Medial Thrust* to some extent. Thus, it makes sense that the *Trunk Lean* strategy is predicted for these subjects.

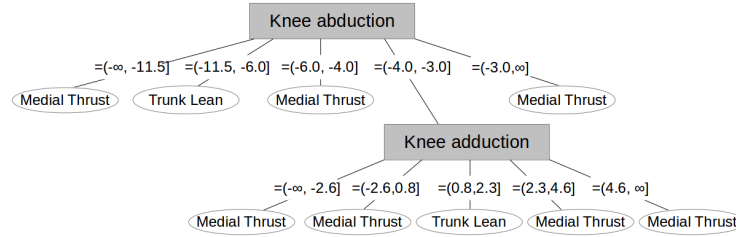


Fig. 1: Decision tree learned for the healthy subjects

```

IF  $-11.5 < \text{Knee abduction} \leq -6.0$  THEN Trunk Lean

IF Toe Out Angle  $> 24.8$  THEN Medial Thrust

IF  $2.8 < \text{Trunk Angle} \leq 4.5$  THEN Medial Thrust

IF KAM Impulse  $> 13.7$  THEN Medial Thrust

IF  $11.0 < \text{KAM Impulse} \leq 13.7$  THEN Trunk Lean

IF  $8.6 < \text{KAM Impulse} \leq 11.0$  THEN Medial Thrust

ELSE Trunk Lean

```

Fig. 2: Rule set learned for the healthy subjects

Arthritic. The best decision tree for this data is shown in Figure 3. The decision tree identifies one range of the tibia angle that results in a *Medial Thrust* prediction. For the other ranges, *Trunk Lean* is predicted. The best rule set for arthritic patients is depicted in Figure 4. The rules that make use of the tibia angle are similar to what appears in the decision tree. The expert said

that rules with the trunk angle feature are less interesting because it is hard to accurately measure one degree differences with portable sensor.

The expert found it interesting that the decision tree and the rule set predict *Trunk Lean* when the tibia angle is larger than six. The tibia angle can be easily measured by positioning a gyroscope on the lower leg, which would potentially allow making predictions outside of a lab setup that uses an expensive camera-based motion capture system.

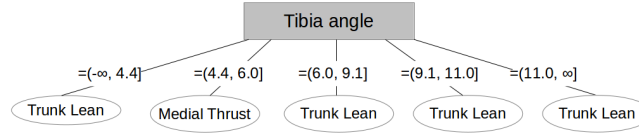


Fig. 3: Decision tree learned for the arthritic subjects

```

IF 6.0 < Tibia angle ≤ 9.1 THEN Trunk Lean
IF 9.1 < Tibia angle ≤ 11.0 THEN Trunk Lean
IF Trunk Angle ≤ 2.0 THEN Trunk Lean
IF 2.0 < Trunk Angle ≤ 3.0 THEN Trunk Lean
ELSE Medial Thrust
  
```

Fig. 4: Rule set learned for the arthritic subjects

4 Conclusions and Future Work

This paper addresses the task of predicting the best gait retraining strategy for an individual with knee osteoarthritis. Several different gait retraining strategies exist, and it is hard to know a priori which strategy will be most suitable for a specific patient. We used machine learning to tackle this problem. We were able to learn reasonably accurate models when training on only healthy individuals or only arthritic individuals. We presented several learned models to a domain expert. She identified several intuitive patterns. Furthermore, she identified a possible location to place a portable on-body sensor that would allow the relevant data to be collected outside of a lab setting that requires expensive equipment.

In the future, we hope to analyze more data collected from patients with knee osteoarthritis, and to explore additional features of an individual's gait. We would also like to analyze data collected using inexpensive, portable sensors for this task.

References

1. Bitton, R.: The economic burden of osteoarthritis. *The american journal of managed care* 15(8) (2009)
2. Fautrel, B., Hilliquin, P., Rozenberg, S., Allaert, F.A., Coste, P., Leclerc, A., Rossignol, M.: Impact of osteoarthritis: results of a nationwide survey of 10,000 patients consulting for oa. *Joint Bone Spine* 72, 235–240 (2005)
3. Fregly, B.J., Reinbolt, J.A., Rooney, K.L., Mitchell, K.H., Chmielewski, T.L.: Design of patient-specific gait modifications for knee osteoarthritis rehabilitation. *IEEE Transactions on Biomedical Engineering* 54, 1687–1695 (September 2007)
4. Fregly, B.J., Rooney, K.L., Reinbolt, J.A.: Predicted gait modifications to reduce the peak knee adduction torque (July 2005)
5. Gerbrands, T., Pisters, M., Vanwanseele, B.: Individual selection of gait retraining strategies is essential to optimally reduce medial knee load during gait. *Clinical Biomechanics* (2014)
6. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: an update. *SIGKDD Explorations* 11 (2009)
7. Lopez, A.D., Murray, C.J.: The global burden of disease, 1990–2020. *Nature Medicine* 4(11), 1241–1243 (November 1998)
8. Mitchell, T.M.: *Machine learning*. McGraw-Hill, Inc., New York, NY, USA, 1 edn. (1997)
9. Miyazaki, T., Wada, M., Kawahara, H., Sato, M., Baba, H., Shimada, S.: Dynamic load at baseline can predict radiographic disease progression in medial compartment knee osteoarthritis. *Annals of the Rheumatic Diseases* 61(7), 617–622 (2002)
10. Mündermann, A., Asay, J.L., Mündermann, L., Andriacchi, T.P.: Implications of increased medio-lateral trunk sway for ambulatory mechanics. *Journal of biomechanics* 41(1), 165–170 (2008)
11. Murphy, L., Schwartz, T.A., Renner, J.B., Koch, G., Kalsbeek, W.D., Jordan, J.M., Helmick, C.G., Tudor, G., Dragomir, A., Luta, G.: Lifetime risk of symptomatic knee osteoarthritis. *Arthritis & Rheumatism (Arthritis Care & Research)* 59(9), 1207–1213 (September 2008)
12. van den Noort, J.C., Schaffers, I., Snijders, J., Harlaar, J.: The effectiveness of voluntary modifications of gait pattern to reduce the knee adduction moment. *Human Movement Science* (2013)
13. Pollo, F.E., Otis, J.C., Backus, S.I., Warren, R.F., Wickiewicz, T.L.: Reduction of medial compartment loads with valgus bracing of the osteoarthritic knee. *The American Journal Of Sports Medicine* 30(3), 414–421 (2002)
14. Quinlan, R.: *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA (1993)
15. Witten, I., Frank, E., Hall, M.: *Data mining: practical machine learning tools and techniques - third edition*. Morgan Kaufmann Publishers (2011)